



C o m m u n i t y E x p e r i e n c e D i s t i l l e d

Data Visualization: a successful design process

A structured design approach to equip you with the knowledge of how to successfully accomplish any data visualization challenge efficiently and effectively

Andy Kirk

[PACKT]
PUBLISHING

Data Visualization: a successful design process

A structured design approach to equip you with the knowledge of how to successfully accomplish any data visualization challenge efficiently and effectively

Andy Kirk



BIRMINGHAM - MUMBAI

Data Visualization: a successful design process

Copyright © 2012 Packt Publishing

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor Packt Publishing, and its dealers and distributors will be held liable for any damages caused or alleged to be caused directly or indirectly by this book.

Packt Publishing has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, Packt Publishing cannot guarantee the accuracy of this information.

First published: December 2012

Production Reference: 1191212

Published by Packt Publishing Ltd.
Livery Place
35 Livery Street
Birmingham B3 2PB, UK.

ISBN 978-1-84969-346-2

www.packtpub.com

Cover Image by Duraid Fatouhi (duraidfatouhi@yahoo.com)

Credits

Authors

Andy Kirk

Project Coordinator

Joel Goveya

Reviewers

Alberto Cairo

Ben Jones

Santiago Ortiz

Jerome Cukier

Proofreader

Chris Brown

Indexer

Tejal Soni

Acquisition Editor

Joanna Finchen

Graphics

Aditi Gajjar

Lead Technical Editor

Shreerang Deshpande

Production Coordinator

Prachali Bhiwandkar

Technical Editor

Dominic Pereira

Cover Work

Prachali Bhiwandkar

About the Author

Andy Kirk is a freelance data visualization design consultant, training provider, and editor of the popular data visualization blog, visualisingdata.com.

After graduating from Lancaster University with a B.Sc. (Hons) degree in Operational Research, he spent over a decade at a number of the UK's largest organizations in a variety of business analysis and information management roles.

Late 2006 provided Andy with a career-changing "eureka" moment through the serendipitous discovery of data visualization and he has passionately pursued this subject ever since, completing an M.A. (with Distinction) at the University of Leeds along the way.

In February 2010, he launched visualisingdata.com with a mission to provide readers with inspiring insights into the contemporary techniques, resources, applications, and best practices around this increasingly popular field. His design consultancy work and training courses extend this ambition, helping organizations of all shapes, sizes, and industries to enhance the analysis and communication of their data to maximize impact.

This book aims to pass on some of the expertise Andy has built up over these years to provide readers with an informative and helpful guide to succeeding in the challenging but exciting world of data visualization design.

Thanks go to my family and friends, but especially to my wonderful wife, Ellie, for her unwavering support, patience, and guidance.

About the Reviewers

Alberto Cairo has taught infographics and data visualization at the University of Miami since January 2012. He is the author of the book *The Functional Art: An Introduction to Information Graphics and Visualization* (Peachpit/Pearson, 2012, <http://www.thefunctionalart.com>). He has been director of infographics at El Mundo online, Spain (2000-2005), professor of infographics and visualization at the University of North Carolina-Chapel Hill (2005-2009), and director of infographics and multimedia at Época magazine, Brazil (2010-2011). In the past decade, he has consulted with media organizations and educational institutions in nearly 20 countries.

Ben Jones is founder of Data Remixed, a website dedicated to exploring and sharing data analysis and data visualization in an engaging way. Ben has a mechanical engineering and business (entrepreneurship) background, and has spent time as a process improvement expert and trainer in Corporate America. Ben specializes in creating interactive data visualizations with Tableau software, and has won a number of Tableau data visualization competitions. This is Ben's first contribution to a book on the subject of data visualization.

I'd like to thank Andy Kirk for selecting me to contribute as a technical reviewer of this book, and my wife Sarah for all the support she gives me in pursuing my passion of the field of data visualization. I'd also like to thank my fellow technical reviewers, from whom I have learned a great deal over the course of the creation of this book.

Santiago Ortiz invents and develops highly innovative and interactive projects for the Web, using self-built frameworks in JavaScript, HTML5, and ActionScript.

He has over more than 10 years of experience working on interactive visualization projects. In 2005, he co-founded Bestiario (<http://bestiario.org>), the first European company specializing in information visualization. Currently, he freelances in the U.S.A. and Europe.

He has presented at events such as VISWEEK, FutureEverything, VizEurope, O'Reilly STRATA, SocialMediaWeek, NYViz, OFFF, and ARS ELECTRONICA.

His projects have been featured in blogs such as ReadWriteWeb, FlowingData, O'REILLY radar, Fast CoDesign, Gizmodo, and The Guardian datablog.

Jerome Cukier is a highly respected Paris-based data visualization consultant with many years of experience as a data analyst and coordinator of data visualization initiatives at the OECD. Jerome specializes in the creation and design

of data visualizations, data analytics, and gamification. His broad portfolio of work is regularly profiled on the leading visualization and design websites and collated on his own site at <http://www.jeromecukier.net>.

www.PacktPub.com

Support files, eBooks, discount offers and more

You might want to visit www.PacktPub.com for support files and downloads related to your book.

Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at www.PacktPub.com and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at service@packtpub.com for more details.

At www.PacktPub.com, you can also read a collection of free technical articles, sign up for a range of free newsletters and receive exclusive discounts and offers on Packt books and eBooks.



<http://PacktLib.PacktPub.com>

Do you need instant solutions to your IT questions? PacktLib is Packt's online digital book library. Here, you can access, read and search across Packt's entire library of books.

Why Subscribe?

- Fully searchable across every book published by Packt
- Copy and paste, print and bookmark content
- On demand and accessible via web browser

Free Access for Packt account holders

If you have an account with Packt at www.PacktPub.com, you can use this to access PacktLib today and view nine entirely free books. Simply use your login credentials for immediate access.

Table of Contents

Preface	1
Chapter 1: The Context of Data Visualization	7
Exploiting the digital age	7
Visualization as a discovery tool	10
The bedrock of visualization knowledge	12
Defining data visualization	16
Visualization skills for the masses	18
The data visualization methodology	19
Visualization design objectives	21
Strive for form and function	21
Justifying the selection of everything we do	22
Creating accessibility through intuitive design	24
Never deceive the receiver	26
Summary	28
Chapter 2: Setting the Purpose and Identifying Key Factors	29
Clarifying the purpose of your project	30
The reason for existing	30
The intended effect	31
Establishing intent – the visualization's function	33
When the function is to explain	33
When the function is to explore	35
When the function is to exhibit data	37
Establishing intent – the visualization's tone	39
Pragmatic and analytical	40
Emotive and abstract	42
Key factors surrounding a visualization project	45

The "eight hats" of data visualization design	48
The initiator	48
The data scientist	49
The journalist	49
The computer scientist	49
The designer	49
The cognitive scientist	50
The communicator	50
The project manager	50
Summary	51
Chapter 3: Demonstrating Editorial Focus and Learning About Your Data	53
The importance of editorial focus	54
Preparing and familiarizing yourself with your data	56
Refining your editorial focus	63
Using visual analysis to find stories	67
An example of finding and telling stories	71
Summary	78
Chapter 4: Conceiving and Reasoning Visualization Design Options	79
Data visualization design is all about choices	80
Some helpful tips	82
The visualization anatomy – data representation	84
Choosing the correct visualization method	84
Considering the physical properties of our data	86
Determining the degree of accuracy in interpretation	87
Creating an appropriate design metaphor	92
Choosing the final solution	94
The visualization anatomy – data presentation	95
The use of color	96
Creating interactivity	107
Annotation	111
Arrangement	115
Summary	117
Chapter 5: Taxonomy of Data Visualization Methods	119
Data visualization methods	120
Choosing the appropriate chart type	121
Comparing categories	122
Dot plot	122
Bar chart (or column chart)	123

Floating bar (or Gantt chart)	123
Pixelated bar chart	124
Histogram	125
Slopegraph (or bumps chart or table chart)	126
Radial chart	126
Glyph chart	127
Sankey diagram	128
Area size chart	129
Small multiples (or trellis chart)	130
Word cloud	131
Assessing hierarchies and part-to-whole relationships	131
Pie chart	131
Stacked bar chart (or stacked column chart)	132
Square pie (or unit chart or waffle chart)	133
Tree map	134
Circle packing diagram	134
Bubble hierarchy	135
Tree hierarchy	136
Showing changes over time	137
Line chart	137
Sparklines	138
Area chart	139
Horizon chart	139
Stacked area chart	140
Stream graph	141
Candlestick chart (or box and whiskers plot, OHLC chart)	142
Barcode chart	142
Flow map	143
Plotting connections and relationships	144
Scatter plot	144
Bubble plot	145
Scatter plot matrix	146
Heatmap (or matrix chart)	146
Parallel sets (or parallel coordinates)	147
Radial network (or chord diagram)	148
Network diagram (or force-directed/node-link network)	149
Mapping geo-spatial data	150
Choropleth map	151
Dot plot map	152
Bubble plot map	152
Isarithmic map (or contour map or topological map)	153
Particle flow map	154
Cartogram	155
Dorling cartogram	156
Network connection map	157
Summary	158

Chapter 6: Constructing and Evaluating Your Design Solution	159
For constructing visualizations, technology matters	159
Visualization software, applications, and programs	161
Charting and statistical analysis tools	161
Programming environments	165
Tools for mapping	167
Other specialist tools	169
The construction process	169
Approaching the finishing line	172
Post-launch evaluation	173
Developing your capabilities	176
Practice, practice, practice!	176
Evaluating the work of others	177
Publishing and sharing your output	178
Immerse yourself into learning about the field	178
Summary	181
Index	183

Preface

Welcome to the craft of data visualization—a multidisciplinary recipe of art, science, math, technology, and many other interesting ingredients. Not too long ago we might have associated charting or graphing data as a specialist or fringe activity—it was something that scientists, engineers, and statisticians did.

Nowadays, the analysis and presentation of data is a mainstream pursuit. Yet, very few of us have been taught how to do these types of tasks well. Taste and instinct normally prove to be reliable guiding principles, but they aren't sufficient alone to effectively and efficiently navigate through all the different challenges we face and the choices we have to make.

This book offers a handy strategy guide to help you approach your data visualization work with greater know-how and increased confidence. It is a practical book structured around a proven methodology that will equip you with the knowledge, skills, and resources required to make sense of data, to find stories, and to tell stories from your data.

It will provide you with a comprehensive framework of concerns, presenting step-by-step all the things you have to think about, advising you when to think about them and guiding you through how to decide what to do about them.

Once you have worked through this book, you will be able to tackle any project—big, small, simple, complex, individual, collaborative, one-off, or regular—with an assurance that you have all the tactics and guidance needed to deliver the best results possible.

What this book covers

Chapter 1, The Context of Data Visualization, provides an introduction to the subject, its value and relevance today, including some foundation understanding around the theoretical and practical basis of data visualization. This chapter introduces the data visualization methodology and the step-by-step approach recommended to achieve effective and efficient designs. We finish off with a discussion about some of the fundamental design objectives that provide a valuable reference for the suitability of the choices we subsequently make.

Chapter 2, Setting the Purpose and Identifying Key Factors, launches the methodology with the first stage, which is concerned with the vital task of identifying the purpose of your visualization – what is its reason for existing and what is its intended effect? We will look closely at the definition of a visualization's function and its tone in order to shape our design decision-making at the earliest possible opportunity. To complete this scoping stage we will identify and assess the impact of other key factors that will have an effect on your project. We will pay particularly close attention to the skills, knowledge, and general capabilities that are necessary to accomplish an effective visualization solution.

Chapter 3, Demonstrating Editorial Focus and Learning About Your Data, looks at the intertwining issues of the data we're working with and the stories we aim to extract and present. We will look at the importance of demonstrating editorial focus around what it is we are trying to say and then work through the most time-consuming aspect of any data visualization project – the preparation of the data. To further cement the learning in this chapter, we will look at an example of how we use visualization methods to find and tell stories.

Chapter 4, Conceiving and Reasoning Visualization Design Options, takes us beyond the vital preparatory and scoping stages of the methodology and towards the design issues involved in establishing an effective visualization solution. This is arguably the focal point of the book as we look to identify all the design options we have to consider and what choices to make. We will work through this stage by forensically analyzing the anatomy of a visualization design, separating our challenge into the complementary dimensions of the representation and presentation of data.

Chapter 5, Taxonomy of Data Visualization Methods, goes hand-in-hand with the previous chapter as it explores the taxonomy of data visualization methods as defined by the primary communication purpose. Within this chapter we will see an organized collection of some of the most common chart types and graphical methods being used that will provide you with a gallery of ideas to apply to your own projects.

Chapter 6, Constructing and Evaluating Your Design Solution, concludes the methodology by focusing on the final tasks involved in constructing your solution. This chapter will outline a selection of the most common and useful software applications and programming environments. It will present some of the key issues to think about when testing, finishing, and launching a design solution as well as the important matter of evaluating the success of your project post-launch. Finally, the book comes to a close by sharing some of the best ways for you to continue to learn, develop, and refine your data visualization design skills.

What you need for this book

As with most skills in life that are worth pursuing, to become a capable data visualization practitioner takes time, patience, and practice.

You don't need to be a gifted polymath to get the most out of this book, but ideally you should have reasonable computer skills (software and programming), have a good basis in mathematics, and statistics in particular, and have a good design instinct.

There are many other facets that will, of course, be advantageous but the most important trait is just having a natural creativity and curiosity to use data as a means of unlocking insights and communicating stories. These will be key to getting the maximum benefit from this text.

You cannot become skilled by reading this book alone, so you need to have a realistic perspective about the journey you are taking and the distance you have made already. However, by applying the techniques presented, then learning and developing from your experiences, you will enjoy a continued and successful process of improvement.

Who this book is for

Regardless of whether you are an experienced visualizer or a rookie just starting out, this book should prove useful for anyone who is serious about wanting to optimize his or her design approach.

The intention of this book is to be something for everyone – you might be coming into data visualization as a designer and want to bolster your data skills, you might be strong analytically but want inspiration for the design side of things, you might have a great nose for a story but don't quite possess the means for handling or executing a data-driven design.

Some of you may never actually fulfill the role of a designer and might have other interests in learning about data visualization. You may be commissioning work or coordinating a project team and want to know how to successfully handle and evaluate a design process.

Hopefully, it will inform and inspire all who wish to get involved in data visualization design work regardless of role or background.

Conventions

In this book, you will find a number of styles of text that distinguish between different kinds of information. Here are some examples of these styles, and an explanation of their meaning.

New terms and **important words** are shown in bold. Words that you see on the screen, in menus or dialog boxes for example, appear in the text like this:

"**Explanatory** data visualization is about conveying information to a reader in a way that is based around a specific and focused narrative."

[ Warnings or important notes appear in a box like this.]

[ Tips and tricks appear like this.]

Reader feedback

Feedback from our readers is always welcome. Let us know what you think about this book – what you liked or may have disliked. Reader feedback is important for us to develop titles that you really get the most out of.

To send us general feedback, simply send an e-mail to feedback@packtpub.com, and mention the book title via the subject of your message.

If there is a topic that you have expertise in and you are interested in either writing or contributing to a book, see our author guide on www.packtpub.com/authors.

Customer support

Now that you are the proud owner of a Packt book, we have a number of things to help you to get the most from your purchase.

Errata

Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you find a mistake in one of our books – maybe a mistake in the text or the code – we would be grateful if you would report this to us. By doing so, you can save other readers from frustration and help us improve subsequent versions of this book. If you find any errata, please report them by visiting <http://www.packtpub.com/support>, selecting your book, clicking on the **errata submission form** link, and entering the details of your errata. Once your errata are verified, your submission will be accepted and the errata will be uploaded on our website, or added to any list of existing errata, under the Errata section of that title. Any existing errata can be viewed by selecting your title from <http://www.packtpub.com/support>.

Piracy

Piracy of copyright material on the Internet is an ongoing problem across all media. At Packt, we take the protection of our copyright and licenses very seriously. If you come across any illegal copies of our works, in any form, on the Internet, please provide us with the location address or website name immediately so that we can pursue a remedy.

Please contact us at copyright@packtpub.com with a link to the suspected pirated material.

We appreciate your help in protecting our authors, and our ability to bring you valuable content.

Questions

You can contact us at questions@packtpub.com if you are having a problem with any aspect of the book, and we will do our best to address it.

1

The Context of Data Visualization

This opening chapter provides an introduction to the subject of data visualization and the intention behind this book.

We start things off with some context about the subject. This will briefly explain why there is such an appetite for data visualization and why it is so relevant in the modern age against the backdrop of enhanced technology, increasing capture and availability of data, and the desire for innovative forms of communication.

After this introduction, we then look at the theoretical basis of data visualization, specifically the importance of understanding visual perception. To help establish a term of reference for the rest of the book, we'll then consider a proposed definition for this subject.

Next, we introduce the data visualization methodology, a recommended approach that forms the core of this book, and discuss its role in supporting an effective and efficient design process.

Finally, we consider some of the fundamental data visualization design objectives. These provide a useful framework for evaluating the suitability of the choices we make along the journey towards an accomplished design solution.

Exploiting the digital age

The following is a quotation from Hal Varian, Google's chief economist (http://www.mckinseyquarterly.com/Hal_Varian_on_how_the_Web_challenges_managers_2286):

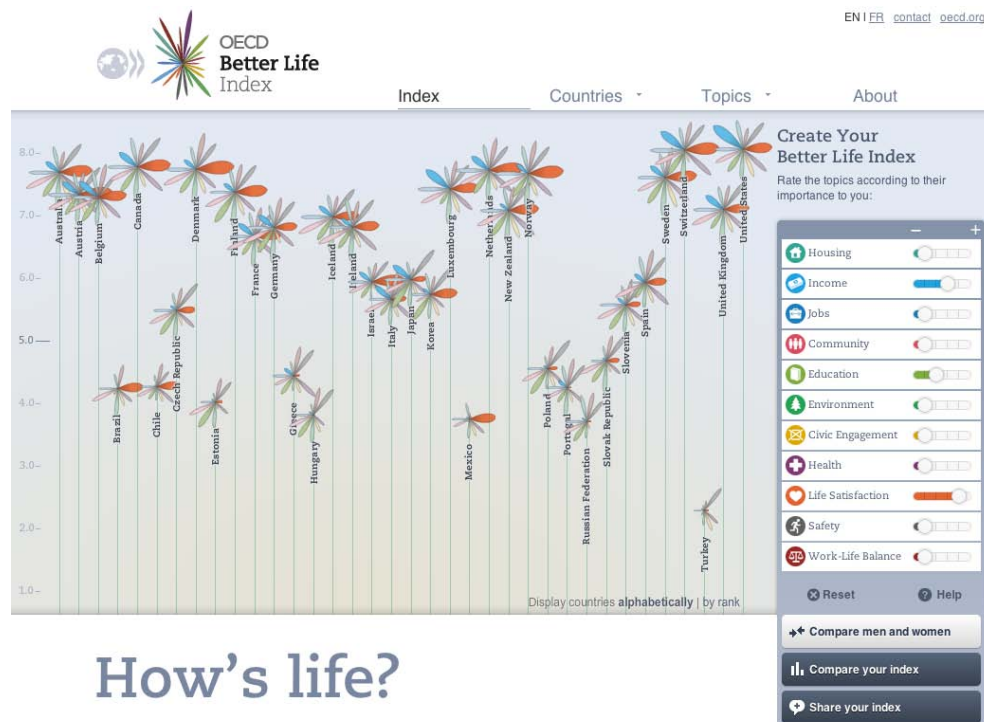
The ability to take data – to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it – that's going to be a hugely important skill in the next decades.

Data visualization is not new; the visual communication of data has been around in various forms for hundreds and arguably thousands of years. Popular methods that still dominate the boardrooms of corporations across the land – the line, bar, and pie charts – originate from the eighteenth century.

What is new is the contemporary appetite for and interest in a subject that has emerged from the fringes and into mainstream consciousness over the past decade.

Catalyzed by powerful new technological capabilities as well as a cultural shift towards greater transparency and accessibility of data, the field has experienced a rapid growth in enthusiastic participation.

Where once the practice of this discipline would have been the preserve of specialist statisticians, engineers, and academics, the globalized field that exists today is a very active, informed, inclusive, and innovative community of practitioners pushing the craft forward in fascinating directions. The following image shows a screenshot of the OECD 'Better Life Index', comparing well-being across different countries. This is just one recent example of an extremely successful visual tool emerging from this field.



How's life?

Image from "OECD Better Life Index" (<http://oecdbetterlifeindex.org>), created by Moritz Stefaner (<http://moritz.stefaner.eu>) in collaboration with Raureif GmbH (<http://raureif.net>)